

2次元シルエットモデルによる人間動作の実時間認識

7D-3

塚本 明利† 李 七雨† 辻 三郎‡
 †イメージ情報科学研究所 ‡大阪大学基礎工学部

1 はじめに

インタラクティブな MMI の実現のためには人間の動作を実時間で認識する技術が必須であり、様々な人物の任意の動作をロバストに認識できる技術が求められている。これまで我々はテクスチャマッピングによって合成した多数のモデル画像を用い、テンプレートマッチングにより顔の向きを実時間で認識するシステムを実現した [1]。本稿では人体の腕や脚の動作を認識する手法として、2次元的なシルエット画像をモデルとしていくつか登録し、それらとのテンプレートマッチングによりモデルの中間状態も含め実時間に認識する手法について述べる。

2 腕、脚の状態認識法

本手法では、人体のシルエット画像と複数のモデル画像とのテンプレートマッチングを行い、両腕および脚の状態を認識する。図1に人体のシルエット画像、図2および図3にそれぞれ上半身、下半身用のモデル画像を示す。個人差や着衣の違いによるシルエット形状の変化に対して安定した認識を行うため、腕の状態認識においては胴体の部分の影響を除去し、腕の部分におけるモデル画像との一致度を算出し、これに基づいて状態認識を行う。

以下では左腕の認識法について述べる。図2(b)の各モデル画像 L_i のうち、腕の部分におけるシルエット画像 S との一致度 $M(L_i)$ は次の式で算出される。

$$M(L_i) = \frac{1}{2} \{D(L_i, U) - D(L_i) + D(U)\} \quad (1)$$

ここで $D(L_i)$ 、 $D(U)$ はそれぞれテンプレートマッチングの結果得られる相違度、 $D(L_i, U)$ はモデル画像間の相違度であり、それぞれ次式で与えられる。

$$D(L_i) = \min_{x,y} \sum_{i,j} |S(x+i, y+j) - L_i(i, j)| \quad (2)$$

Real-Time Gesture Recognition using 2-Dimensional Silhouette Models
 Akitoshi Tsukamoto†, Chil-Woo Lee† and Saburo Tsujii‡
 †Laboratories of Image Information Science and Technology
 1-4-2 Shinsenri-Higashimachi, Toyonaka, Osaka, 565 Japan
 ‡Faculty of Engineering Science, Osaka University

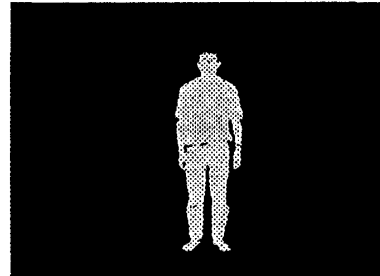


図1: 人体のシルエット画像

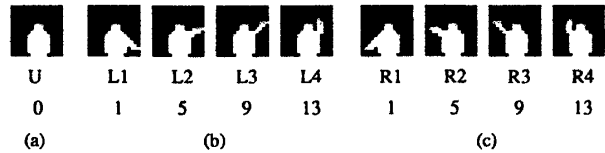


図2: 上半身用モデル画像と状態番号

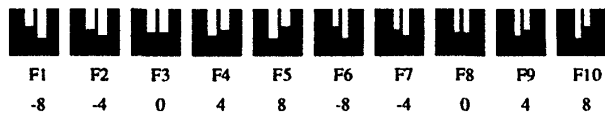


図3: 下半身用モデル画像と状態番号

$$D(L_i, U) = \sum_{i,j} |L_i(i, j) - U(i, j)| \quad (3)$$

腕の状態は、一致度がもっとも高いモデル画像を選択し、それと隣接するモデル画像との相違度を比較することにより認識する。図2(b)のうち最小相違度を与えたモデル画像を L_j とし、閾値を Th とするとき、以下の判定により状態番号 L を得る。

$$L = \begin{cases} Num(L_j) + 2 & M(L_j) = M(L_{j+1}) \\ Num(L_j) + 1 & M(L_j) > M(L_{j+1}) > Th \\ Num(L_j) & M(L_{j\pm 1}) < Th \\ Num(L_j) - 1 & M(L_j) > M(L_{j-1}) > Th \\ Num(L_j) - 2 & M(L_j) = M(L_{j-1}) \end{cases} \quad (4)$$

これは、腕の状態が連続的に変化する際に隣接するモデル画像との一致度が図4のように変化することによる。なお、 $Num(L_j)$ は L_j の状態番号を表す。

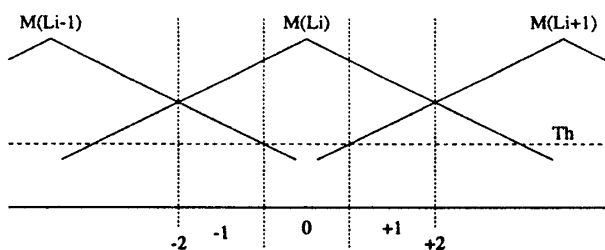


図 4: モデル画像との一致度と中間状態番号

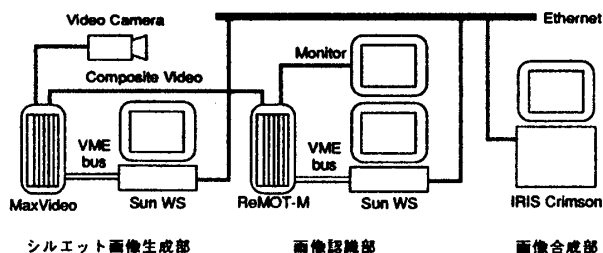


図 5: 実時間動作認識システム構成図

3 システム構成

図 5 に実時間動作認識システムの構成を示す。このシステムは以下の各モジュールから構成されている。

シルエット画像生成部 Datacube 社の MaxVideo システムを用い、入力画像と背景画像との差分を閾値処理してシルエット画像を作成する。得られた画像はビデオ信号により出力し、またシルエット領域の位置情報を画像認識部に伝送する。

画像認識部 ReMOT-M[2] にシルエット画像を入力してテンプレートマッチングを行い、結果の相違度に基づいてジェスチャ認識を行う。

画像合成部 得られた状態番号を受けとり、それに基づいて簡単な人体のアニメーションを作成し、認識結果の表示を行う。

なお、背景およびビデオカメラの位置、ズームングなどは固定とし、衣服などの影響を低減させるため暗幕の前で動作を行うよう撮影環境を準備した。

4 実験結果

暗幕の前で足踏み、その場走り、腕の回転や腕を上上げるなどの動作を行い、そのときの腕や足の状態の認識を行った。図 6 にテンプレートマッチングの結果を示す。図中の白枠はマッチング位置を示すものであり、右下の数値はそれぞれ右腕、左腕、および足の状態番号、および相違度を示す。

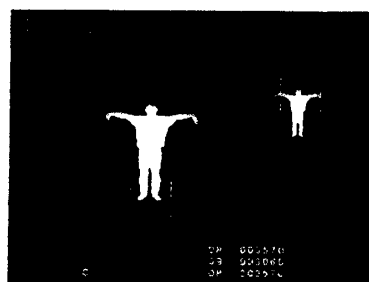


図 6: テンプレートマッチング結果

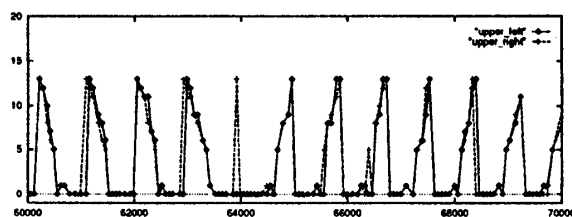


図 7: 両腕の状態認識結果

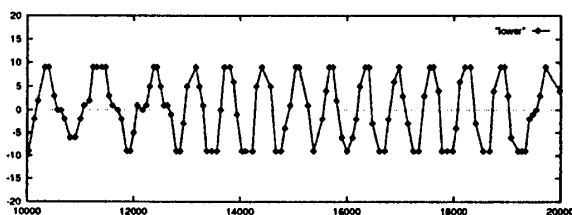


図 8: 脚の状態認識結果

図 7 および図 8 に実験の結果得られた状態番号の経時変化を示す。同図より、腕の回転や足踏みなどの動作が認識されていることがわかる。

5 おわりに

本稿ではシルエット画像とのテンプレートマッチングによって実時間にジェスチャ認識を行う手法について述べた。本手法は個人差や衣服の違いに対し安定で、かつモデル画像の中間状態を推定することも可能であり、実時間で処理可能であることからインタラクティブシステムへの情報入力手段として有効である。

参考文献

- [1] 塚本, 李, 廣田, 辻: “複数のモデル画像による顔の向きの実時間推定”, 第 10 回ヒューマン・インタフェース・シンポジウム論文集, pp. 85-92 (1994).
- [2] 廣田, 李, 徐, 辻: “相関法を用いた実時間物体追跡システムの製作(2)”, 情処 48 全大, No. 5M-4, pp. 73-74 (1994).